

# A Framework for an Ego-centered and Time-aware Visualization of Relations in Arbitrary Data Repositories

Florian Reitz

Dept. of Databases and Information Systems  
University of Trier, Germany  
reitzf@uni-trier.de

**Abstract.** Understanding constellations in large data collections has become a common task. One obstacle a user has to overcome is the internal complexity of these repositories. For example, extracting connected data from a normalized relational database requires knowledge of the table structure which might not be available for the casual user. In this paper we present a visualization framework which presents the collection as a set of entities and relations (on the data level). Using rating functions, we divide large relation networks into small graphs which resemble ego-centered networks. These graphs are connected so the user can browse from one to another. To further assist the user, we present two views which embed information on the evolution of the relations into the graphs. Each view emphasizes another aspect of temporal development. The framework can be adapted to any repository by a flexible data interface and a graph configuration file. We present some first web-based applications including a visualization of the DBLP data set. We use the DBLP visualization to evaluate our approach.

## 1 Introduction

The amount of data we create and store increases every day. The result is a growing number of large and possibly complex repositories. Much effort has been invested in defining potent query languages like SQL and XQuery and there are many fast and reliable algorithms to apply them. However, these techniques do little to hide the internal structure of the collection. For example, a highly normalized relational database might require several table joins to gather all information which belongs to one entry. While an expert can use information on the internal structure to speed up queries, a casual novice might easily be overstrained. This is a problem because the user wastes valuable time needed for the actual task on understanding the data organization.

A common way to hide internal complexity is to present the collection as a set of entities which are connected by various types of relations. Consider the DLBP bibliographic data set<sup>1</sup>. It contains meta data for 1.3 million publications. *Authors* are one type of entity that can be found in the collection. They are connected by the *coauthor* relation if they cooperated on at least one publication. There are several approaches to use this model in applications. Some prefer a textual representation [1] [9] while others

<sup>1</sup> dblp.uni-trier.de

provide a more complex visualization[10] [15]. Relations also provide a context for a given entity and thereby support exploring the data set.

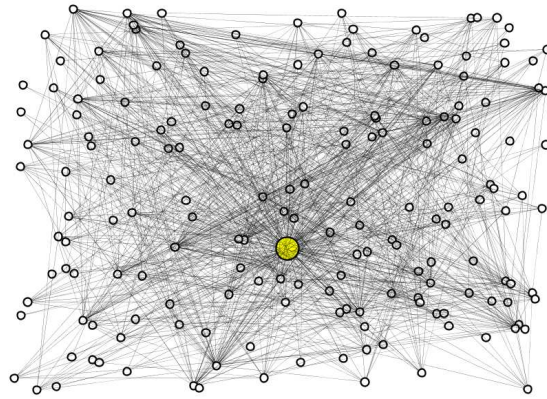


Fig. 1: *Adam's* complete neighborhood in the DBLP coauthor graph. *Adam's* node is painted in yellow

The networks which are defined by relations in real-world data sets can be very large and dense. The DBLP coauthor network consists of 750,000 authors and 2.5 million relations. Suppose we want to examine the publication behavior of author *Adam*. Even finding *Adam* in this network is cumbersome. Figure 1 shows a node-link drawing of *Adam's* direct neighborhood in the coauthor network. It still contains 178 nodes and 1154 edges (177 *Adam* - coauthor + 977 coauthor - coauthor). The drawing hides the fact that some relations are stronger than others. For example, *Adam* has cooperated 25 times with *Bob*, but only 1 time with Jack, so the tie to *Bob* is much stronger. It also does not reveal at which time the relations were formed and how they have evolved over time.

In this paper we show how relation networks can be divided into small connected graphs. In Section 2 we describe how filtering and rating can reduce the neighborhood graphs. We obtain simple graph drawings which resemble ego-centered networks [5]. These drawings can be enriched with information on how a relation has evolved. In Section 3 we present two views which cover different aspects of evolution. One emphasizes *when* a relation was influenced while the other focuses on *how strong* this influence was. A single graph is of little use if it is not linked with others. In Section 4 we show how the graphs are embedded into a framework which allows browsing the graphs. In Section 5 we present examples of visualized relations in DBLP and the German language Wikipedia. We conclude this paper with a two-part evaluation of our approach.

## 2 Relation graphs

A common approach to visualize graphs is the node-link diagram. Entities are represented by nodes. If two entities are in relation, their nodes are connected by an edge. In most real-world data sets there are a number of different relations. The networks defined by these relations can become very large and do not allow comprehensible node-link drawings. To reduce their size, we generate a set of ego-centered drawings.

### 2.1 Inverted Ego Graphs

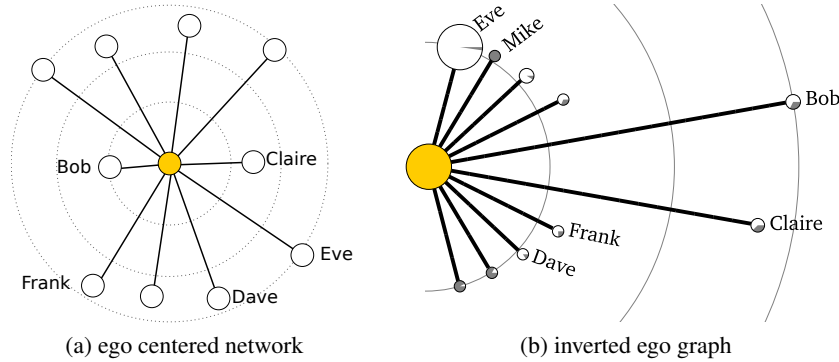


Fig. 2: *Adam's* coauthor network as ego graphs. *Adam's* node is painted in yellow.

In a first step we split the relation networks into small graphs. Each graph represents the direct proximity of one entity. We call this entity the *ego* of the drawing. In the coauthor example, we obtain 750,000 graphs, one for each author. Figure 1 shows that this approach will be virtually useless if the drawing violates several aesthetic criteria which have been identified for this type of drawing [2] [7]. Above all, the number of edge crossings is problematic here. To improve the comprehensibility of the split graphs we only include the most relevant neighbors. For most relations there is a straightforward definition of relevance. In the coauthor example, we can use the number of joint publications. If we put the node which represents ego in the center of the drawing and place the selected neighbors (the *alters*) in a distance according to their relevance, we will obtain an ego-centered network [5]. We do not draw edges between alters to avoid edge crossings. Figure 2a shows *Adam's* ego-centered network. We have included the ten most relevant alters. *Bob* is placed closest because the relation with him is the strongest. *Eve* is the least relevant of the included authors.

In Section 3 we will use the edges to display information on how the relation has evolved over time. In many cases, the most relevant alter is of special interest and we can expect that the development of this relation is more complex than others. However, the edge which represents this relation is the shortest in the drawing and thereby provides

the least space for additional information. To increase the available space, we introduce the **inverted ego graph**. We place the most relevant alter at maximum distance to the ego while less important alters are positioned closer. To further increase the edge length, we place the ego and the most relevant alter at opposite sides of the drawing. This contradicts the *close related entities are placed close to each other* metaphor, but it provides sufficient space. Figure 2b shows *Adam’s* inverted ego graph.

## 2.2 Drawing Details







type	parameters	examples
none	-	
solid	<i>one color</i>	
fraction	$d \in [0, 1]$ , <i>one color</i>	
pie	Map (color $\rightarrow$ double)	
timecolor	list of colors	
presence	list of Booleans, <i>one color</i>	

Table 1: The available node fillings. Parameters in italics are optional.

Ego-centered networks originate from the field of qualitative network analysis. Aside from the relevance value, these drawings often contain additional information which might be important for a user who interprets them. We can add information by modifying size, shape and filling of the nodes. For example, in Figure 2b we use the size of the alters to show the total number of publications. Obviously, *Eve* is by far the most active author. However, she has a low relevance because only a small part of her publications were joint work with *Adam*. We use the node filling to show this fraction. *Mike* for example published all work with *Adam*. Together with the low number of publications, this suggests that he is a PhD student rather than an established researcher. Both ego and alters represent persons so we use the same node shape to draw both.

Table 1 shows the six types of node fillings and the parameters they require. `none` leaves the nodes empty while `solid` fills them with a specified color. `fraction` requires a value  $d \in [0, 1]$  which defines how much of the node area is filled. `pie` is a generalization of `fraction` where multiple color segments can be used. In the `timecolor` filling, all segments have the same size and the whole area is covered. The `presence` filling divides the area into equal-sized parts. A list of Boolean values defines which parts are filled and which are left white. In Section 5 we will show examples of all types.

### 3 Time Views

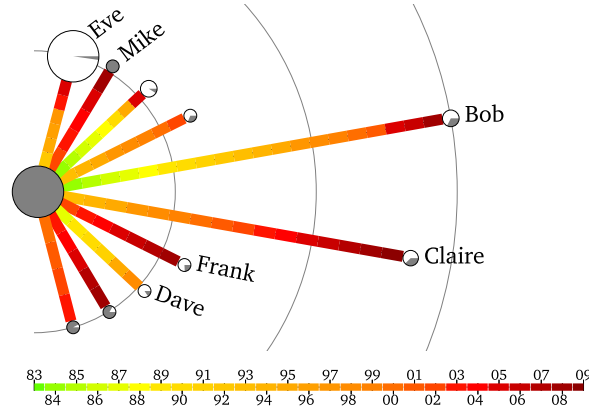


Fig. 3: The time-color view for *Adam's* coauthors

Not all relations have evolved in the same way and a user might be interested to see these differences. For example, if we look for *Adam's* long-term partners, we want to discriminate them from coauthors with a short but intense cooperation. First, we divide the time frame, i.e., the interval between the oldest and the youngest time stamp in the data set into equal-sized periods. The data set must contain sufficient information to determine how strong a relation was at each period. In the DBLP example, the time stamps represent the year of publication. The oldest paper is from 1936, the newest from 2010. We use 75 periods, each covering a single year. The strength of a relation in a single period is the number of joint papers in that year. This information is available in the data set. There are periods in which the relation becomes stronger and periods in which it remains unchanged. We do not consider relations which become less intense over time.

We use the edge between ego and alter to show the development in different periods. There are two aspects we must consider: *when* was the relation influenced and *how strong* was this influence. For each aspect we have implemented a view, i.e., a way to modify the edge drawings. The positions of the nodes are the same for both views so we can easily switch between the two to see both aspects.

#### 3.1 Time-Color View

The time-color view emphasizes the moment in which a relation was influenced. We assign a color to each period which expresses its position in the time frame. There is no linear order of colors but most people accept the purple-blue-green-yellow-orange-red sequence for this purpose [14]. We assign purple and blue tones to periods from the beginning of the time frame. Red tones indicate recent periods. For each period which

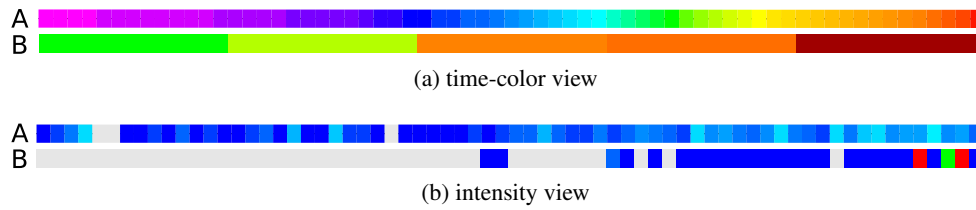


Fig. 4: Two edges from time-color view and intensity view respectively with different development

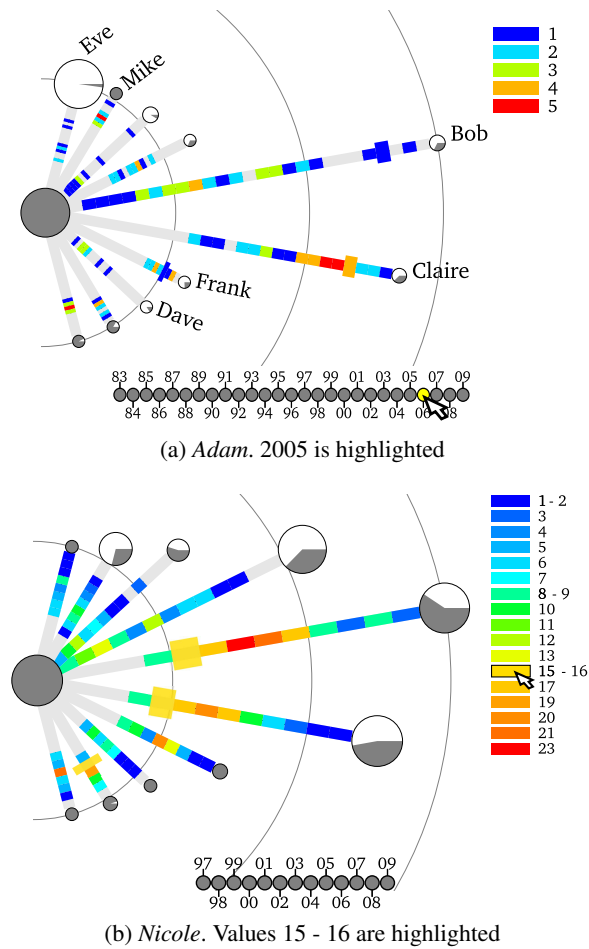


Fig. 5: The inverted coauthor ego network of Adam and Nicole in the intensity view

affected the relation, regardless how strong, we add a colored segment to the respective edge. The segment color matches the represented period. Periods with no influence have no representation. The segments are ordered by time, the oldest close to the ego and the newest close to the alter. Figure 3 shows *Adam*'s coauthors in the time-color view. At the bottom, there is a color bar which shows the time-color mapping. Green and yellow segments refer to the 1980s while reddish segments represent the years after 2000. We can see that the relation between *Adam* and *Bob* has evolved over a long time while the cooperation with *Claire* started later. The cooperation with *Dave* ended in the late 1990s.

The segment size depends on the number of periods which are relevant for the respective relation and the available edge length. Figure 4a shows two edges with different content. While edge A covers many consecutive periods, B shows two distinct phases of development marked by an abrupt color change. In a first version of this view, we modified a segment length based on the strength of influence. The segments of important periods became longer compared to those of less relevant periods. Early user feedback (see Section 6.1) showed that this additional distortion confused the viewers. We gave up on this feature in favor of the intensity view.

The number of periods is limited by the number of colors a human can discriminate. In an ideal environment, we can distinguish more than one million colors[8] but the user feedback showed that even the 75 DBLP periods can become problematic. Especially, matching segment colors with the bottom bar was reported to be difficult. We use linking and brushing to compensate. Whenever the mouse cursor moves over a segment, all segments of the same color are painted in double stroke. The bottom bar shows only those segments which are relevant for at least one relation.

### 3.2 Intensity View

The intensity view emphasizes how strong a relation was influenced during a period. We use colors to show the strength of development. Periods with little influence are painted in blue while more important ones are presented in reddish tones. Unlike the time-color view, each edge contains segments for all periods. Like before we leave out periods which are not relevant for any relation. If a period had no influence on a relation, the segment is painted in white so it appears as a gap. Because the number of segments is the same for all edges, one level of distortion is eliminated. Now we can use the position of the segments to guess the period. Cleveland and McGill [4] showed that humans can perceive positions much better than colors. This compensates for the smaller segments. Figure 4b shows two edges from the intensity view. While edge A shows a homogeneous but weak development with only three gaps, edge B represents a younger relation with a strong recent development.

Figure 5a shows *Adam*'s coauthor graph in the intensity view. The strength of development in a period is equivalent to the number of joint publications in this year. Although *Bob* and *Claire* have similar importance values, we can clearly see the differences in development. It is also visible that the last cooperation between *Adam* and *Dave* happened some time ago. An additional legend at the right side of the graph maps colors and values. In Figure 5a there are only five different development values. In Figure 5a all segments of the year 2005 are highlighted by brushing over the bottom bar. In

Figure 5b we see the coauthors of *Nicole* who published 23 papers in cooperation with a single author in a single year. If the domain becomes too large, the framework maps intervals to a color instead of single values. In this drawing, all segments with 15 or 16 joint papers are highlighted.

## 4 Framework Components

The framework consists of three major components, the user interface, the graph generator and the interface with the underlying data.

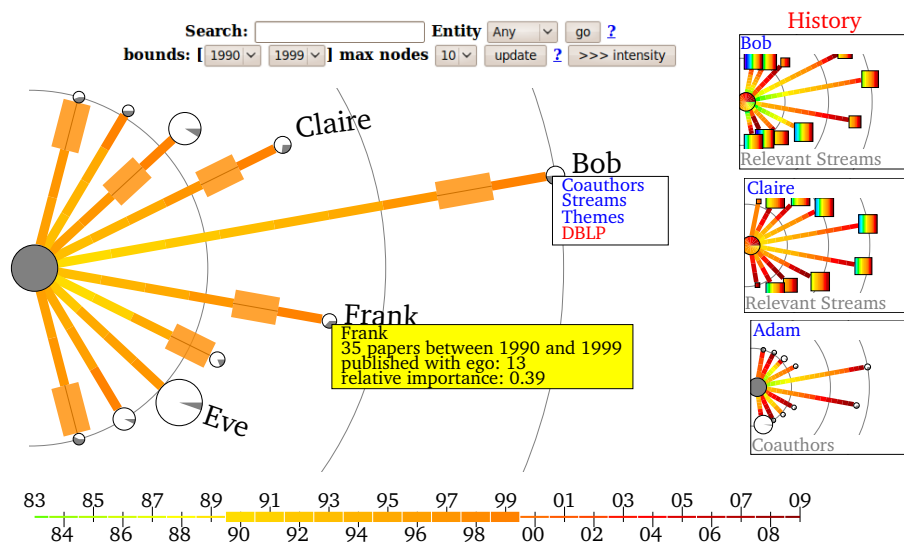


Fig. 6: The user interface showing an example of *Adam's* coauthor graph in the time-color view.

The front end is composed of the graph drawing and some additional elements which assist the user. We use the SVG (Scalable Vector Graphics) format for the drawings. Vector graphics are smaller than pixel graphics and can be zoomed without loss of quality. There are a number of frameworks which allow SVG rendering in applications and many web browsers support a sufficient part of the standard. Figure 6 shows an example of a web-based front end displayed by the Firefox 3.0 browser. The component positions are a suggestion and can be changed if needed.

The most important front end function is to link the graphs. We can click on a node and a new graph is created where this node is the ego. In Figure 6 there is more than one possible relation for a person type ego. A connection menu (Bob) lists the available options. In the example, there is also an external link to *Bob's* DBLP author page. In theory, SVG graphics can provide smooth changeovers between the graphs where nodes



move to their new positions and new ones are faded in. However, only few browsers support SVG animation. We expect this to improve in the future.

There are two other ways to switch to a new graph. The head menu provides a search field which can be used to find a new ego if the node is not contained in the current graph. The menu also contains control elements for the time lens feature. The time lens allows the user to define which periods should be considered for displaying the graphs and computing the relevance function. In Figure 6 we see *Adam* as if he only published in the 1990s. The bottom bar shows which periods are relevant here. If we compare the relevance values with those in Figure 3 we clearly see some differences. The head menu also contains a control to modify the maximum number of alters. Often, it is useful to go back to a previous graph. The history bar shows thumbnails of up to four old requests. We can click the thumbs to enlarge the drawings again. In some situations it is useful to have textual representation of data. All nodes have a tooltip menu which appears when the cursor moves over it. In the next section, we will show how they are defined. Like all visual effects, including those we discussed in Sections 3.1 and 3.2, tooltips are created by JavaScript functions which are embedded into the SVG files. No additional command processing is necessary.

There is no limit to the complexity of the underlying data source. In most cases we need sophisticated information extraction strategies and hand tuned queries to achieve acceptable performance. The data which is necessary to draw an entity or a relation might be scattered over the whole data set. Only an expert can provide a fast interface with these repositories by implementing a given Java interface. An important part of this step is to actually define the entities and relations. The expert can utilize caching strategies and pre calculations if necessary and enforce privacy policies for example by making entities anonymous. Other framework components request data from the interface by giving a pre-defined key and a data type. A similar technique is used to define the rating function for the relations.

## 5 Example Applications

To demonstrate the framework, we deployed two applications on data sets with different characteristics: DBLP and the German language Wikipedia.

### 5.1 DBLP

The coauthor graph drawings we used in this paper are taken from this application. In this section we give some background information and additional examples for drawing details. This visualization is available online<sup>2</sup> and generated some user feedback. The DBLP data set is available as an XML file which lists all records. The size and the structure of this file make efficient queries impossible so we import the document into a relational database. During this step, we pre calculate some frequently used values. We take additional information on journals and conferences from html pages on the DBLP web server.

---

<sup>2</sup> <http://dblpvis.uni-trier.de/>

**Entities:** We extract three types of entities (number on December 8, 2009): *person* (760,277), *word* (36,150) and *stream* (3480). Stream is the term DBLP uses to refer to conferences and journals. Each author with at least one publication listed in DBLP is represented by a *person* entity. DBLP uses person names as identifiers. If there are two homonymous authors, a numeric suffix is appended to the name. If a person uses different names the data set contains entries to match them. The search engine is customized to consider this additional data as well. The *word* entities are derived from the publication titles. We remove stop words and invert inflections.

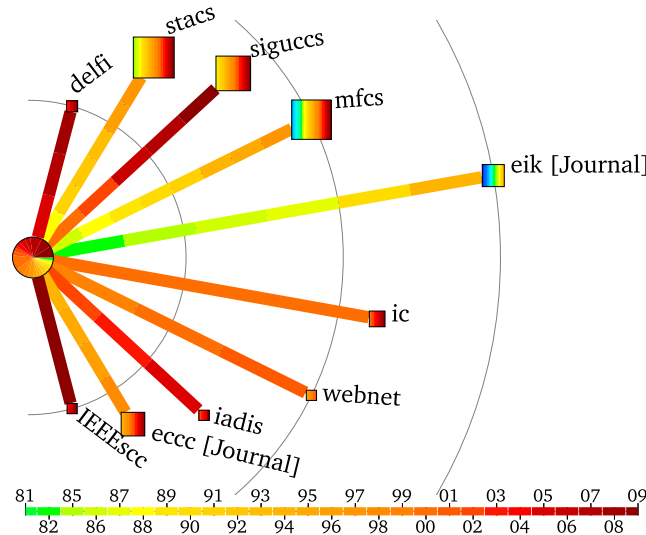


Fig. 7: *Petra*'s relations to conferences and journals in the time-color view

**Relations:** Figure 7 shows the relations between *Petra* and the streams which accepted her papers. The relevance for this relation is defined by the number of accepted papers. The edge coloration shows that *Petra* stopped publishing at *eik*, *mfcs* and *stacs* in the 1990s. To better understand the reasons for this, we use a time-color filling for the stream nodes. For each year a conference or a journal was active (held a venue or published an issue) we add a segment in the respective color. If a stream is old some colors will not be shown in the bottom bar. Missing red segments show that *eik* was not continued after 1994 which explains why there are no further publications. However, *mfcs* and *stacs* were continued after *Petra*'s last publications so there must be other reasons which the data set does not reveal. With the exception of *SIGUCCS* all stream nodes with recent publications are small. This means they do not have many years of activity and are rather young. The pie filling of the ego node shows when *Petra* was active. The more papers she published in a year the larger the respective segment.

Figure 8 shows which themes were popular at the *Journal of Symbolic Logic (JSYML)*. *JSYML* is the stream with the most active years in DBLP. We use the words extracted from the publication titles. They are a poor replacement for actual keyword lists but pro-

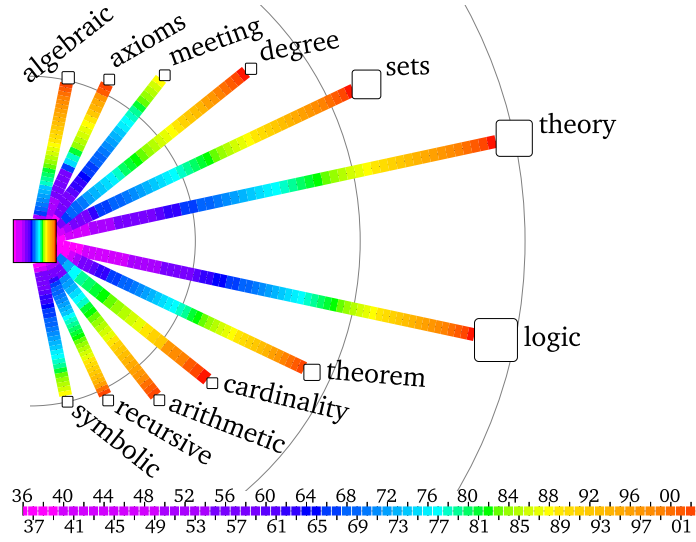


Fig. 8: Related themes of *JSYML* in the time-color view

vide acceptable results. Based on the titles from a single year we apply *term frequency - inverse document frequency* (tf-idf)[13] to sort out nondescriptive words like *system*. Kuhn and Wattenhofer[9] used a similar approach to get thematic descriptions of conferences. The ego node has a time-color filling. Red segments are missing here because *JSYML* was not continued after 2003. We can see different categories of themes. *meeting* and *symbolic* were used from the beginning but do not appear later. *cardinality* was not used at the beginning and *theory* and *logic* appeared at all time. Note that *meeting* and *cardinality* have a similar importance value although they developed differently. Figure 9 shows the opposite relation. We see the streams related to the word *query* in the intensive view. tf-idf returns values that are difficult to interpret but the higher the value the stronger the influence. We use presence filling to show when a stream started. For each relevant period we add a segment. If the alter was active in this period the segment would be painted blue otherwise it would be painted white. The *webdb* conference for example was established late and therefore could not use any keywords in early years. We can see that *icdt* is a biennial conference because only every second segment is colored.

There are other relations as well but without additional drawing details. The most important one is a relation between streams which is rated by comparing communities and themes. We also provide a relation between author and relevant themes.

## 5.2 Wikipedia

As another application we considered relations in the German language Wikipedia. This data set differs from DBLP in size and density of the relation networks. There are 4.6 million authors (including unregistered) and 2.7 million pages. A page is an article, a

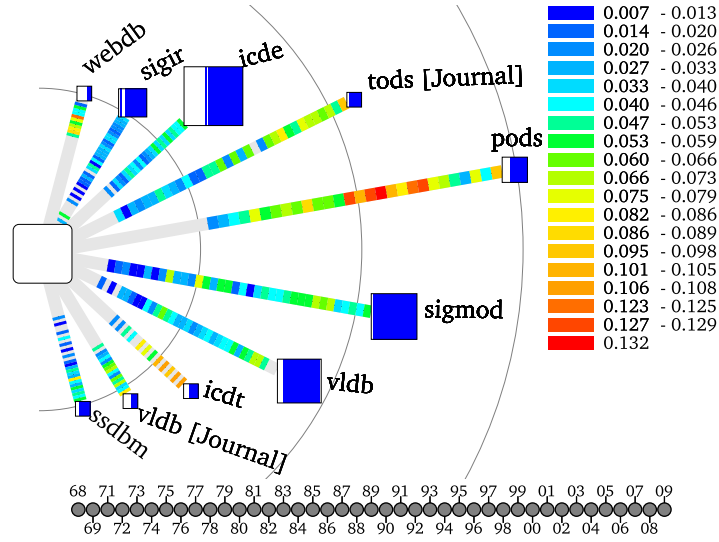


Fig. 9: Related streams for the keyword *query*

user page or an administrative page. While an average DBLP author contributed to 2.89 streams, a Wikipedia author modified on average 12.59 articles. This requires more sophisticated caching and pre-calculation strategies for the data interface. There are also differences in the time frame. While DBLP only contains 75 time stamps (1936-2010) the Wikipedia dump we considered was changed on more than 3000 days since 2001. Via the data interface, we define a period as a month and map the respective time stamps to it.

Figure 10 shows the relation between *Gil* and the articles she modified. *Gil* is administrator so we paint the ego node as a circle. Otherwise, we would have used a rounded rectangle for registered and a triangle for unregistered users. We use presence filling to show in which months an article was changed. The color we use for the presence filling of the alters shows the type of page.

## 6 Evaluation

The evaluation of our approach consists of a field study in which we observed people using the visual interface and a controlled experiment in the laboratory. We used the DBLP data set (Section 5.1) in either case.

### 6.1 Basic Field Study

In January 2009, we launched a web application using the DBLP data set. For a period of 320 days, we logged which type of graph was requested and which settings were used. To get an approximate mapping of requests and users, we also logged a hash of the session ID. Sessions with a very high number of requests (most probably web

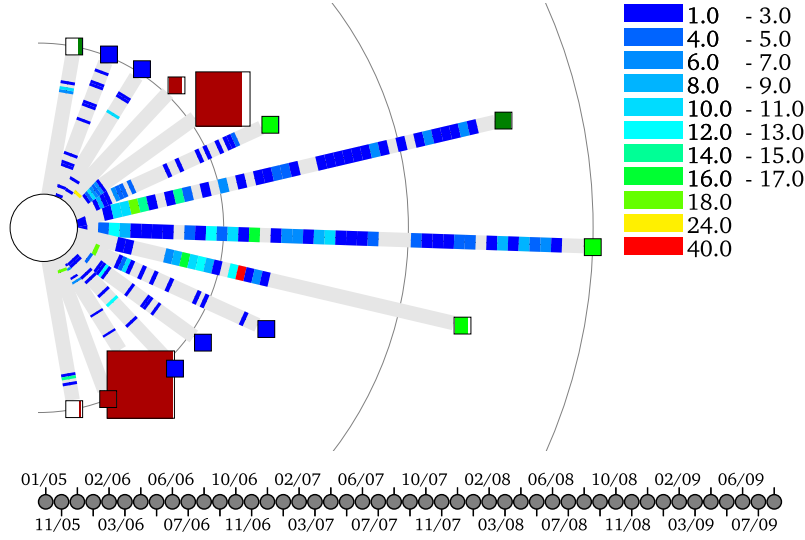


Fig. 10: *Gil's* Wikipedia article contributions. Green colors: user pages. Red colors: administration pages, Blue colors: articles.

bots) were excluded. During the observation, there were 42,068 sessions with a total of 107,683 requests. Most sessions terminated after the first request but 1277 times the user viewed more than ten graphs. The mean length of these long sessions is 30.5 which indicates that the application was actually used for browsing. The time lens was used more often when the session length increased while the use of the search engine dropped. The intensity view was added later so there is no significant data which view was favored.

We also received direct feedback. In Section 3.1 we already described the remarks on the time-color view and the resulting modifications. The tenor was that the users considered the program to be useful but only after they gathered some experiences with interpreting the drawings. Especially the drawing details seem to pose the risk of confusion and misinterpretations. In general, there were two groups of users. One requested additional information while the other favored simple drawings.

## 6.2 Task-based Study

To get more direct information of the usefulness of our approach we conducted a study in the laboratory.

**Participants:** Two female and eight male undergraduate students aged 22 to 29 participated in the study. All rated themselves as regular computer and web users. Eight participants stated to know the DBLP data set but only one had advanced experience. Nobody was experienced with this framework and the associated visualizations.

**Setup:** After a short introduction on the visual interface and DBLP, we gave the participants time to familiarize themselves with the application. Then we asked them to complete three groups of tasks ( $G_1$ ,  $G_2$  and  $G_3$ ) with three problems each.

- $G_1$  Tasks could be solved by analyzing a single graph.
- $G_2$  Tasks required using a specific feature (for example time lens) or a specific view.
- $G_3$  *Open tasks*. We asked participants to explore the neighborhood of a given entity. Noteworthy constellations should be reported.

Because no participant was experienced in the application domain, we had to limit the complexity of the tasks in  $G_3$ . For  $G_1$  and  $G_2$  we expected short answers to specific questions. For  $G_3$  we were mainly interested in how the users applied their experience from  $G_1$  and  $G_2$ . We randomized the order of tasks within each group to counter learning effects. We observed the participants and logged which graphs they requested and which parameter they used. The tasks had to be completed in 30 minutes. At the end of the session, the participants were asked to answer open end questions on what they liked and didn't like on the framework.

**Task Results:** Eight participants solved all  $G_1$  tasks and six all  $G_2$  tasks. Nobody failed more than one task in a group. For each task we asked if completing it had been easy, slightly difficult or difficult. In 55 out of 60 cases, the task was rated easy. The remaining cases were considered to be slightly difficult. Eight participants were able to find interesting constellations in  $G_3$ . The others reported only uninteresting details or actual misinterpretations. Nine participants stated that the lack of knowledge on the DBLP data set was the major problem. All were positive to find more if given additional time. By observing the users and analyzing the log files we found out that:

- All Tasks (or parts of task) which required to tell the relevance of an entity were correctly solved.
- At first, the users preferred the time-color view. A task in  $G_2$  forced them to use the intensity view. After that, this view was preferred.
- In  $G_2$  half of the users browsed to a new graph using the search function rather than the connection menus. With growing experience this behavior changed. In  $G_3$  all participants used the connection menus if possible.
- Not understanding drawing details was the major cause of errors. While there was no problem with the fraction filling, many users misinterpreted the fillings of the stream nodes (see Figure 7). Eight participants reported that they were confused because the filling used colors which did not appear in the bottom bar.
- Tooltips were used more often than we expected. All participants tried to validate their interpretation of the drawing with textual information if possible.
- The time lens was considered an important part of the application. It was used in many cases, even if it did not contribute to the solution.

In general, the comments were positive. The users liked the small and clear graph drawings and the idea of presenting a rating by the length of an edge. Only one person criticized the fact that *close* related alters are placed *far* away from the ego. Like in the field study feedback, learning to use the application was a major issue. Four participants explicitly stated that they had to learn how to use the application first. There also was a significant learning effect. Unlike the feedback from the field study, the participants requested more textual information. Four persons explicitly mentioned text integrated in the drawings. For example, the relevance value should be visible next to the nodes. Two

proposed an additional view which should contain the information in tabular form. This supports our finding that tooltips are frequently used. The tooltips were consistently mentioned as a positive aspect.

Four participants stated that the information density was too high. But, like in the field study feedback, all users posted ideas on what information should be added. Three participants proposed additional drawing details like patterns or more types of node shapes. This shows that a graph definition has to be done very carefully with respect to the user group. The definer must not give in on the wish to integrate as much information as possible but has to make reasonable selections.

## 7 Related Work

Aggregated approaches show the temporal data in a single drawing. LifeLines[12] visualizes a person's disease pattern. For each condition there is a timeline, i.e., a horizontal bar along a time axis. Coloration and thickness of these bars change to show the status of the condition at different times. TimeRadarTrees[3] uses a radial drawing to show how several entities are related with each other. Unlike our approach, there is no ego and all relations between the visible nodes are displayed. Instead of node-link, it uses colored segments which fill a circle with multiple layers. The approach is limited to small graphs but supports hierarchical nesting to compensate this. Segments at the perimeter represent recent events while those near the center represent old influences. The intensity view is similar to this drawing but has a much lower information density. ConfSearch[9] searches DBLP for relations between conferences, authors and keywords. The related entities of an ego are presented as a rated list with additional information. ConfSearch does not show the evolution of relations, but we adopted some of the rating functions used for the examples in Section 5.1.

Many systems combine different types of visualizations. Paper Lense [10] and Facet Lense [11] provide bar charts, textual result lists and nested node drawings to show entities in faceted data sets. Among others, data can be plotted against time, like the number of an author's publications by year. Both tools provide extensive filtering and sorting functions. The DB-Browser[15] features similar views including simple graph drawings. PaperLense and DB-Browser visualize DBLP data. Both provide aggregated information like the number of joint papers for two given authors.

Not all approaches use the entity and relation abstraction. The ThemeRiver[6] application shows how the frequency of a term in a set of documents changes over time. The results for multiple terms are presented as a plot where one axis shows the time and the other axis the frequency.

## 8 Conclusion And Future Work

In this paper, we presented a framework that can visualize relations in any given data set. Large and complex relation networks are reduced to small graphs. The drawings of these graphs resemble ego-centered networks and convey information on when and how strong a relation evolved. The drawings are part of a visual interface which supports the user in understanding the data and links the graphs with each other. Experiments

from an online application and the results of a basic usability study indicated that our approach is useful though it poses the risk of generating graphs which overstrain the user. Future work will have to address the problem of how difficult it is to understand the drawing details. These studies will also have to include other drawing details.

There is no clear definition of an ego-centered graph in the literature. Many definitions allow additional edges between the alters or even nodes that are in no direct connection with the ego. The decision to include a node or an edge is usually based on complex strategies or the intuition of the person who generates the graph. It is unclear whether the inverted ego network could be extended this way and whether this would benefit the user.

## References

1. H. Bast, A. Chitea, F. M. Suchanek, and I. Weber. ESTER: efficient search on text, entities, and relations. In *SIGIR*, pages 671–678. ACM, 2007.
2. C. Bennett, J. Ryall, L. Spalteholz, and A. Gooch. The Aesthetics of Graph Visualization. In *Computational Aesthetics*, pages 57–64. Eurographics Association, 2007.
3. M. Burch and S. Diehl. TimeRadarTrees: Visualizing Dynamic Compound Digraphs. *Computer Graphics Forum*, 27(2):823–830, 2008.
4. W. S. Cleveland and R. McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491 – 500, 1986.
5. L. C. Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291 – 304, 1982.
6. S. Havre, E. G. Hetzler, P. Whitney, and L. T. Nowell. ThemeRiver: Visualizing Thematic Changes in Large Document Collections. *IEEE Trans. Vis. Comput. Graph.*, 8(1):9–20, 2002.
7. Helen C. Purchase. Metrics for Graph Drawing Aesthetics. *Journal of Visual Languages and Computing*, 13(5):501 – 516, 2002.
8. P. K. Kaiser and R. M. Boynton. *Human Color Vision*. Optical Society of America, Washington, DC, 2nd edition, 1996.
9. M. Kuhn and R. Wattenhofer. The Layered World of Scientific Conferences. In *APWeb LNCS 4976*, pages 81–92, 2008.
10. B. Lee, M. Czerwinski, G. G. Robertson, and B. B. Bederson. Understanding research trends in conferences using paperLens. In *CHI Extended Abstracts*, pages 1969–1972. ACM, 2005.
11. B. Lee, G. Smith, G. G. Robertson, M. Czerwinski, and D. S. Tan. FacetLens: exposing trends and relationships to support sensemaking within faceted datasets. In *CHI*, pages 1293–1302. ACM, 2009.
12. C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. LifeLines: Visualizing Personal Histories. In *CHI*, pages 221–227, 1996.
13. G. Salton and C. Buckley. Term-Weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manage.*, 24(5):513–523, 1988.
14. C. Ware. Color Sequences for Univariate Maps: Theory, Experiments, and Principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, 1988.
15. A. Weber, P. Reuther, B. Walter, M. Ley, and S. Klink. Multi-Layered Browsing and Visualisation for Digital Libraries. In *ECDL*, volume 4172 of *LNCS*, pages 520–523. Springer, 2006.